

# ARIS: Integrating Multi-source Data for Research in Andrology

<sup>1</sup>Teun Timmers Ph.D., <sup>1,2</sup>Frank Pierik M.Sc., <sup>1</sup>Marc Steenbergen B.Sc.,  
<sup>1</sup>Huibert Stam B.Sc., <sup>1</sup>Astrid M van Ginneken M.D. Ph.D., <sup>1,3</sup>Erik M van Mulligen Ph.D.,  
<sup>2</sup>Robertus FA Weber M.D. Ph.D  
<sup>1</sup>Dept. of Medical Informatics, <sup>2</sup>Dept. of Andrology, <sup>3</sup>University Hospital Dijkzigt,  
Erasmus University Rotterdam, The Netherlands

*Although the concept of distributed systems for the storage of patient data is more and more commonly accepted, for some considerable time yet most patient data will be stored in centralized rather than departmental systems. An important advantage of storage in a central system is hospital-wide access to much of the patient data. Disadvantages are however that these data cannot be reviewed through one user interface, and that the structure of the data does not lend itself to exploitation for other purposes.*

*We describe the implementation of an Andrology Research Information System in which these data are integrated in a well-structured database facilitating multiple views on the patient data through a graphical user interface, and clinical research, quality control and summary reports. The data can be analyzed directly using the Hermes workstation. In this way the strengths of the centralized system are combined with those of the dedicated ARIS system.*

## INTRODUCTION

Infertility is a problem with which 10-15% of all couples are confronted while trying to conceive. In about one third of all couples seeking professional help for fertility problems show indications that male subfertility is the reason for the inability to conceive. In addition to patient care, the Department of Andrology is involved with clinical research in this area of health care. For example, methods are developed and evaluated for improving sperm quality, and for predicting the chances to conceive. Data are collected on the patients visiting the Andrology outpatient clinic in the BAZIS/ZIS Hospital Information System (HIS) during patient visits and from the result of laboratory tests. In addition, data are collected on paper and in PC databases.

This situation poses two problems. First, a complete overview of all data on a particular patient is not available through one user interface. Second, when data are extracted from the HIS and the other data sources, these data are available as separate files with a structure that makes them unsuitable for statistical analysis.

Tools have already been developed at the Department of Medical Informatics to solve both problems.

First, several years ago the development was started of a Computer Patient Record (CPR) which enables the detailed recording of events, actions and problems and their links [1,2]. This has resulted in a database design for CPRs in internal medicine and a Cardiology outpatient clinic. In the latter the CPR is currently operational.

Second, the Hermes workstation provides a client-server based software architecture integrating a range of modules supporting clinical data analysis [3]. The Hermes architecture also provides the basis of the latest implementation of the CPR. Using these tools, we have implemented the Andrology Research Information System (ARIS), which solves the problems described. The project was in three stages:

1. integration of the data coming from the various sources into a well-defined database, introducing an improved structure of the data. The existing CPR database design was extended to include the data collected at the Department of Andrology. During off-line import into the database the data are validated;
2. building a graphical user-interface to the database, which enables multiple views on the data. The generic software components of the CPR and the Hermes client-server kernel were used;
3. making the database available for analysis with the tools for clinical research in the Hermes workstation.

We will discuss these stages in detail in the following.

## INTEGRATION OF THE SOURCE DATA

We have three kinds of source data:

1. data stored in a specialized subsystem of the BAZIS/ZIS HIS, created with the REGIST dedicated software package. Data contained in this source are those pertaining to history, past history, semen analysis, diagnoses, therapy, and follow-up of therapy. These data are entered by the physicians during patient visits and by laboratory personnel.
2. data from the Clinical Chemistry laboratory. These data are also available through subsystems of the

BAZIS/ZIS but they have to be extracted separately. Included are the results of hormone tests such as Testosterone, FSH and SHBG.

3. data collected at the Department of Andrology, e.g. a detailed record on Intra-Uterine In-seminations and their results, and newly developed tests for infertility. These data have been collected on paper and in PC databases.

These data are extracted from the systems on a regular basis. For the purposes of clinical research it is advantageous to have a snapshot database. The data extractions from the systems are flat ASCII files, which must be validated. The data have different formats and may use inconsistent codes. Validation and coding conversions are done when the data are imported into the ARIS database. In total the extractions consist of 7 files with a total number of approximately 175 attributes.

The structure of the extraction files also poses several other problems. In the semen analysis file for example, there are items describing the result of semen analysis. A semen analysis can be performed in several ways (in the ejaculate or in serum), and for several reasons (testing its quality, or feasibility for IUI). These qualifications are stored in separate attributes, thus making simple queries impossible. Another problem is the relative lack of explicit time stamps for many events, excluding accurate recording of chronology. These problems are solved by the suitable structuring of the ARIS database, described below.

## ARIS DATABASE STRUCTURE

We used the philosophy of the Computer Patient Record model developed at our Department, which aims at the explicit recording of actions, events and problems, and the links between them [1,2]. At the moment, designs for a CPR in internal medicine and a cardiology outpatient clinic have been made; the latter system is currently in use.

The CPR model allows the detailed recording of:

- the *problem* of the patient: the condition for which the patient is being evaluated and/or is under treatment;
- *actions*: e.g., a history, lab result or prescription;
- *events*: an event is a set of one or more actions that belong together.

This CPR model therefore allows problem-oriented record-keeping and recording of evolution of insight. The latter is achieved by explicit time-stamping of when an insight occurred, when the insight was confirmed or rejected, and since when the insight was valid. In addition to this so-called *macro-structure*, a

*micro-structure* of the CPR is being developed, enabling the detailed formal presentation of the findings that are otherwise written as free text in the history and physical exam [4]. In the source data only little free text was included so these possibilities have not yet been further pursued.

Using the CPR model to design the ARIS database structure meant, on the one hand, breaking up the original data files into parts that constitute the above actions and events, and on the other, reconstructing - insofar as possible - when these events and actions occurred. Normalization was done to the level of the Boyce-Codd normal form [5]. This resulted in a database consisting of 38 relations and 273 attributes, as opposed to the original 7 files containing 175 items. For example, the original history file contains 109 attributes. This file is expanded in ARIS into 22 relations. For each record in the history file, 81 records are generated in ARIS. The database currently contains data on 7000 patients.

An important conclusion from this exercise is that the CPR model, designed originally for internal medicine and a cardiology outpatient clinic, is sufficiently general and powerful for easy adaptation to another specialist area.

The database has been implemented in InterBase (Borland International, Inc). Using its API, we have installed it as a server for both the ARIS graphical user interface and for the Hermes system components that support clinical research. All software runs on HP9000 hardware under UNIX and is written in C, using the X11 OSF/MOTIF Library. For client-server communication the Berkeley socket mechanism is used.

## THE GRAPHICAL USER INTERFACE

The graphical user interface is a client to the ARIS database. We have used the generic software components built by the CPR development group to create the ARIS user interface. Its windows are tied together in the user interface by means of the User Interface Engine [6]. An essential feature of the User Interface Engine is that it works on a high level of abstraction. That is, the behavior and contents of the windows are described by definition files.

The windows in the user interface have been built with the UIM/X graphical interface builder. Currently the ARIS user interface contains 19 windows, of which 9 are part of the general CPR, and 10 are specific for Andrology.

To summarize, to build the ARIS GUI only the specific Andrology windows had to be created, and the definition files mentioned above had to be written.

Patient Profile						
Name	Birthdate	Sex	Patient id	Couple #	Date profile	
		Male / Female	1378779	529	23 Jan 1995	

Diagnoses		
17 Mei 1990	Afsl. of liesbr	voorlopig
17 Mei 1990	Inf. primair	onbekend
17 Mei 1990	Oligoasth. sp.	voorlopig

Past history		
Bof:	onbekend	
Cryptorchisme:	nee	
Hernia Inguinalis:	nee	
Prostatitis:	nee	

Visits		
17 Mei 1990	Eigen afdeling	Eerste consu

Test results	
6 Apr 1990	Semenanalyse
18 Mei 1990	C.K.C.L.
30 Jul 1990	Semenanalyse
21 Sep 1990	I.U.I.

Medications	

Figure 1. The ARIS Patient Profile Window, showing a summary report on the patient. The Test results window shows tests originally stored in different systems. They are now seen in chronological order, showing their context.

Using the generic software components (the User Interface Engine and the communication with the database), we have been able to create the ARIS user interface together with the definitions of the callback SQL queries within only a few man months. It should be noted that currently ARIS is not used for the entry of patient data, although this feature is available in the software. It has been disabled because data collection takes place at several sites in the Hospital.

In Figure 1, the Patient Profile Window is shown. This window presents a summary report of the patient data. In the *Test results* section the integration of the data is seen most clearly: nearly every entry comes from a different source. It would be much more complicated and would provide less insight if a similar chronology of events had to be reconstructed from the various sources, although this is possible to some degree in the HIS. By double-clicking on one of the test results, another window is popped up, showing all data available on that particular test.

## CLINICAL RESEARCH

One of the main goals for ARIS is to facilitate clinical research in the field of Andrology. Once the data are available in the ARIS database, the Hermes workstation components can also access these data.

Hermes is based on a client-server model with additional brokering for location of services and it provides a software integration architecture with which legacy systems can be integrated using an encapsulation mechanism [3]. Currently it comprises clients and services for: (i) user-friendly database query construction, (ii) graphical inspection and presentation of data, and (iii) statistical analyses. The same architecture is used for client-server communication in ARIS. In fact, ARIS can be seen as just another service in the Hermes environment.

Essential for Hermes is the concept of a *meta-datamodel*, described in detail elsewhere [7]. The datamodel is available as a service in the Hermes environment. It contains metadata such as codings for coded attributes, pretty names for these attributes, and the types of the data. Support of clinical research is greatly facilitated by an additional *medical datamodel*, also present in the datamodel service. This is an extended datamodel which presents the data in a medically meaningful way. The user can choose to use this datamodel in the Hermes client with which database queries can be constructed in a user-friendly way. Figure 2 shows part of the medical datamodel.

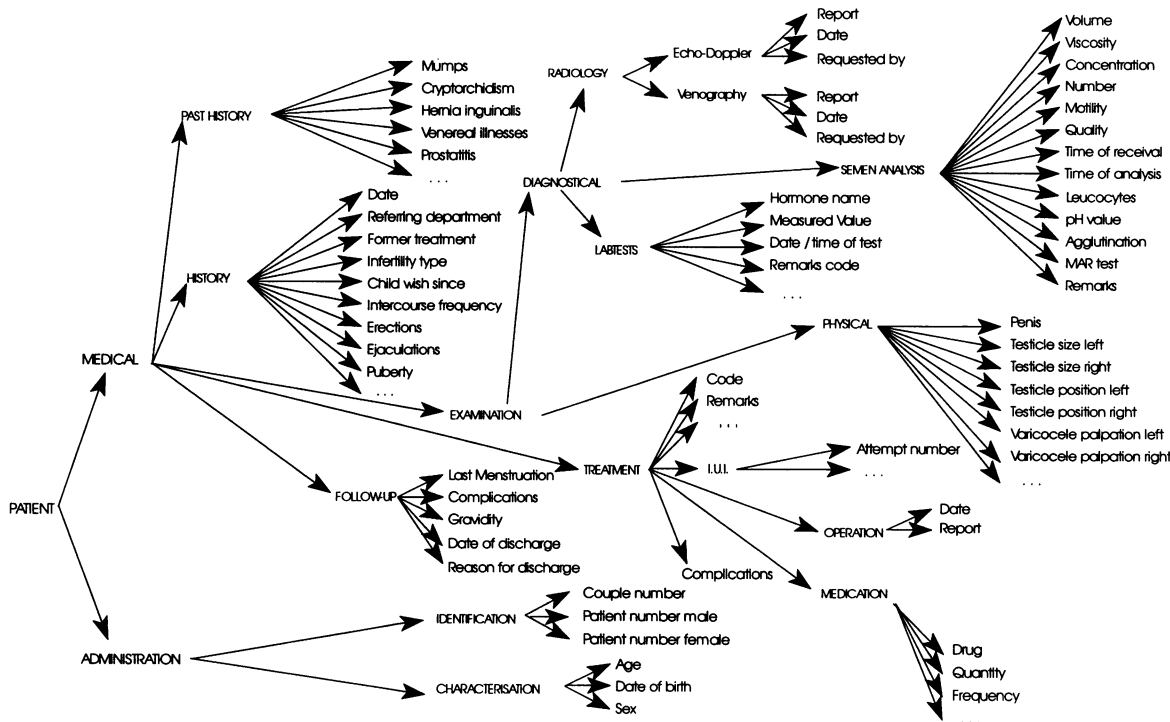


Figure 2. Part of the medical datamodel as available in the Hermes workstation. From this display the user can create a query to the ARIS database. The items are presented with their medical names and in their medical context, rather than with their attribute names. The datamodel generates the proper relation names and keys automatically.

## CONCLUSIONS

For some considerable time yet a great deal of patient data will be collected in legacy Hospital Information Systems. The important advantage is hospital-wide availability of these data for patient care. Complementary to this, there is a need for the possibility of reviewing all data pertaining to a patient through a common user interface, allowing multiple views. Another vital need is to exploit the data for clinical research, quality assessment and summary reports. ARIS satisfies both needs and can be used as a complement to the existing data entry systems. The development of the ARIS system has been immensely facilitated by the use of generic software components for the ARIS user interface and a CPR model already developed for internal medicine and cardiology. This model developed at our Department appeared to be of sufficient general application to allow for the fast development of ARIS.

## ACKNOWLEDGMENTS

M. Steenbergen built ARIS to complete requirements for his M.Sc. thesis. M. de Wilde and R. Cornet contributed significantly to the implementation.

## References

1. Van Ginneken AM, Stam H, Duisterhout JS. A powerful macro model for the Computer Patient Record. In: Ozbolt J (ed.), Proc. 18th SCAMC, 1994, 496-500.
2. Van Ginneken AM, Stam H. Can one patient record accommodate the diversity of specialized care? In: Proc. 19th SCAMC, 1995.
3. Van Mulligen EM, Timmers T, Van Bommel JH. A new architecture for integration of heterogeneous software components. *Meth Inf Med* 1993; 32:292-301.
4. Moorman PW, van Ginneken AM, van der Lei J, et al. A model for structured data entry based on explicit descriptonal knowledge. *Meth Inf Med*; 33(5): 454-463.
5. Elmasri R, Navathe SB. *Fundamentals of Data-base Systems*. 1992. The Benjamin / Cummings Publishing Company Inc. Redwood City, Cal.
6. Stam H. Design and implementation of a generic User Interface Engine. Internal Report, Dept. of Medical Informatics, Erasmus University Rotterdam, The Netherlands.
7. Timmers T, van Mulligen EM, van den Heuvel F. Integrating clinical databases in a medical workstation using knowledge-based modelling. In: Lun KC et al. (eds.), Proc. Medinfo '92, North Holland, Amsterdam, 478-482.